

# DPMs for the Density Estimation in a Dynamic Nonlinear Modeling: Application to GPS Positioning in Urban Canyons.

Asma Rabaoui\*, Nicolas Viandier<sup>†</sup>, Juliette Marais<sup>†</sup> and Emmanuel Duflos\*<sup>‡</sup>

\* LAGIS UMR CNRS 8146, Ecole Centrale de Lille, Villeneuve d'Ascq, France.

<sup>†</sup> INRETS, LEOST, Villeneuve d'Ascq, France.

<sup>‡</sup> Team-Project SequeL, INRIA Lille Nord-Europe, Villeneuve d'Ascq, France.

e-mail: asma.rabaoui@ec-lille.fr

## Abstract

In Global Positioning System (GPS), classical localization algorithms assume, when the signal is received from the satellite in line-of-sight (LOS) environment, that the pseudorange error distribution is Gaussian. Such assumption is in some way very restrictive since a random error in the pseudorange measure with an unknown distribution form is always induced in constrained environments especially in urban canyons due to multipath/masking effects. In order to ensure high accuracy positioning a good estimation of the observation error in these cases is required. To address this, an attractive flexible Bayesian nonparametric noise model based on Dirichlet process mixtures (DPM) is introduced. Since the considered positioning problem involves elements of non-Gaussianity and non-linearity and besides it should be processed on-line, the suitability of the proposed modeling scheme in a joint state/parameter estimation problem is handled by an efficient particle filter called Rao-Blackwellized particle filter (RBPF).

Our approach is illustrated on a data analysis task dealing with joint estimation of vehicles positions and pseudorange errors in a GNSS based localization context where the GPS information may be inaccurate because of multipath/masking effects.

## Index Terms

Global Positioning System (GPS), navigation, urban canyon, pseudorange errors, density estimation, nonparametric Bayesian Methods, sequential Monte Carlo methods, Rao-Blackwellized particle filter.

## I. INTRODUCTION

GPS is a radio navigation system which relies on radio-frequency signals emitted by a constellation of satellites. Consequently, it can be easily distorted by the measuring system and propagation impairments (atmospheric layers, multipath, diffraction and mask phenomena). Today, Global Navigation Satellite Systems (GNSS) have penetrated the transport field through a variety of applications such as monitoring of containers, fleet management or car navigation. These applications do not necessarily request high availability, integrity and accuracy of the positioning system. However, for safety applications (such as guidance of autonomous vehicles), performances require to be much more stringent.

Unfortunately, most of all these transport applications are mainly used in dense urban environments. Even with modern GPS receivers, high positioning accuracy is only achieved in LOS conditions when the signal is received directly without any reflexion. Since the signals delivered by sensors are easily blurred because of hard external conditions in urban canyons, they may deliver totally erroneous measurements, especially when the signal reaches the receiver after interaction with several objects/obstacles. In urban areas, some obstacles (buildings, trees, etc) can be modeled using a simulator and thus some of the errors can be predicted or modeled. However, some other obstacles (cars, pedestrians, etc) can appear suddenly and can induce a random error in the pseudorange measure. In order to ensure high accuracy positioning, a good estimation of the observation error in these cases is required.

The incoming signal is in general the sum of the direct signal and several delayed replica. However, non-LOS signals, which are signals received after reflections on the surrounding obstacles, frequently occur in dense environments. In these environments, most of the time no direct signals can reach the receiver and thus localization accuracy is severely degraded due to high propagation time delays.

In our paper, we consider positioning accuracy degradation problem when we don not receive the direct signal but just reflected replica. In this case, the receiver tracks only reflected signals. Such phenomena make the pseudorange error distribution becomes a non-centered Gaussian distribution because of additional errors on pseudorange estimation. As a consequence, classical localization methods assuming that state and observation noises are zero mean Gaussian are not efficient anymore and can severely impair positioning accuracy.

### A. Previous Work

Literature focusing on techniques for localization performance enhancement in constrained environments is abundant. The most spread rely on multi-sensors approaches [1], [2], which goals are to compensate the lack of GNSS performances by adding other sensors (odometer, inertial measurement unit, etc). However multi-sensors methods show some drawbacks and in particular the system cost and complexity.

In [2], [3], [4], the proposed approaches allow to adapt the error model in the filtering process to the reception condition of each satellite signal. This modeling permits to handle the overall reception process which switches between some preselected measurement models. The switching process permits to define three reception states (LOS, non-LOS and no reception cases). The principle idea of such modeling scheme is to define an indicator variable which governs the behavior of the statistical model and allow for switching from one model to another. An interesting way of defining the statistical structure of this indicator variable has been developed in [2].

Switching state-space models, also called jump state-space models, have been widely studied in the literature.

In [5], [6] a Jump Markov Linear System (JMLS) has been proposed for the case where the considered state-space model is conditionally linear and Gaussian. The estimation problem was handled using finite mixture modeling (Interacting Multiple Model (IMM) [7], [8] and Generalized Pseudo-Bayes (GPB) [5] algorithms). For non-linear and non-Gaussian cases, efficient sequential Monte Carlo algorithms have been applied [9], [10], [11], [12].

In [3] the proposed Jump Markov System (JMS) shows quite good results in terms of accuracy and stability in the context of GPS positioning. These acceptable performances are owed to the adaptation of the error model according to the reception state. However, the proposed approach for error estimation using the Expectation-Maximization (EM) algorithm has some limits especially related to slow convergence. Moreover, the non-LOS case is modeled by a finite Gaussian mixture and the number of Gaussian components is fixed. Consequently, the mixture model can not represent the true measurement error.

In this paper, in order to limit costs, we have chosen to work only with GNSS signals, no additive sensors will be used. For a better accuracy, we propose a new statistical filtering method based on a better definition (and use) of the observation noise for each satellite signal. In order to ensure high accuracy positioning a good estimation of the observation error in such cases is required.

### *B. The Motivating Problem*

In this paper, the state evolution equation governing the position of a land vehicle is a dynamic non-linear model and the measurement equation (observation model) is nonlinear as well with additive observation errors. Within this setting, a joint estimation problem of hidden states and noise probability densities will be considered. Often, the observations which are the GPS measurements arrive sequentially in time and one is interested in performing the joint estimation on-line.

Note that in a dynamic nonlinear model with additive observation errors, it is usual to assume that errors are normally distributed or can be approximated by a finite Gaussian mixture. This can cause problems when there are for example outlying errors leading to many modes in the density distribution form whose variability over time induces poor inference about the model parameters. As we are considering in this paper the case where multipath propagation affects severely GPS signals qualities, it is our intention therefore to model the observation errors using a highly flexible family of density functions. It is important to notice that mixture modeling is considered as a successful and widely used density estimation method, capable of representing the phenomena that underlie many real-world datasets. However, the main difficulty in mixture analysis is how to choose the number of mixture components. Model selection methods in general treat the number of components as an unknown constant and set its value based on the observed data. Such an approach lacks flexibility, since in practice we often need to model the possibility that new observations come from as yet unseen components.

We propose to use a flexible density modeling based on a Bayesian nonparametric approach involving an infinite mixture model. The Dirichlet Process Mixture (DPM) model, studied in Bayesian nonparametrics [13], [14], is the most used approach for nonparametric density estimation. The construction of probability measures on the space of distribution functions has received considerable attention in the Bayesian nonparametrics literature [15]. A rich literature exists on its theoretical properties, and it has been used in a variety of problems [16], [17], [18]. Despite this activity, only a few work has been conducted on how to tune the hyperparameters of a nonparametric mixture model. Most of the time, in real-world application contexts, the observed data can help in defining a prior for some

of the hyperparameters in the considered Bayesian model. Computing every unknown hyperparameter with pure sampling is not the most effective approach in all cases since this can be practically difficult and time consuming.

Our objective is to estimate jointly the hidden states  $\{x_t; t \in \mathbb{N}\}$  and the error density parameters that pertain to the observation noise vectors  $\{w_t; t \in \mathbb{N}\}$ . These density parameters are the latent variables  $\{\theta_t; t \in \mathbb{N}\}$  to be estimated using a nonparametric Bayesian approach. For on-line Bayesian filtering, the particle filter permits to compute the posterior distribution  $p(x_t, \theta_t | z_{1:t})$  recursively over time. In this case, we have to take into consideration the fact that computing everything with pure sampling can be computationally expensive. One solution consists in considering the augmented hidden state vector  $\{Y_t = (x_t, \theta_t); t \in \mathbb{N}\}$ . By the partition of the state-space  $Y_t$  into two sub-spaces drawn by  $\theta_t$  and  $x_t$  (for all  $t \geq 1$ ), improved formulations of the particle filter can be applied. Rao-Blackwellization [19], [20] is one way of improving the efficiency of the particle filter. In the case where it is possible to evaluate some of the filtering equations analytically and the others with Monte Carlo sampling, Rao-Blackwellization will lead to estimators with less variance than what could be obtained with pure Monte Carlo sampling [21]

The idea is to partition the state vector so that one component of the partition is a conditionally nonlinear Gaussian state-space model; for this component one can work out the solution analytically and use the Extended Kalman filter to compute  $p(x_{1:t} | \theta_{1:t}, z_{1:t})$ . The particle filter is then used only for the nonlinear non-Gaussian portion of the state-space to compute  $p(\theta_{1:t} | z_{1:t})$ . In this way, we can say that the hard part of the problem will be reduced into the computation of the posterior distribution  $p(\theta_{1:t} | z_{1:t})$ .

### C. Contributions and paper organization

This paper considers the problem of density estimation from a Bayesian nonparametric viewpoint, using a hierarchical mixture model. A typical choice is to assume the unknown density as a mixture of a parametric family with a discrete random probability as a mixing distribution. This random probability is induced by the Dirichlet Process and the hierarchical model is the Dirichlet Process Mixture model. By using DPM, it is important to mention that no matter what additive errors distributions are involved we are confident that our family of densities will be able to capture the right shape and hence statistical inference for the parameters of interest will be improved and reliable. A drawback of using a model based on the Dirichlet process is that it is infinite dimensional and therefore inference will be complicated. However, recent innovations in sampling algorithms within infinite dimensional frameworks has made considerable progress in recent years to such an extent that it is now possible to perform exact inference without the need to set up arbitrary approximations.

The flexibility of DPM models grows when we assume one more step in the hierarchy, i.e. when the the DPM hyperparameters are random. Our aim using this family of densities is first to be able to capture the right shape of the noise probability density functions (pdfs) and then to improve the statistical inference for the parameters of interest. Therefore, we will show how to tune the DPM hyperparameters in a flexible way. The posterior distribution some hyperparameters will be estimated as part of the Gibbs sampler, and some other parameters will be chosen carefully in a data-adaptive way for a better fitting of the data distribution shape.

To sum up, this paper proposes several contributions. The first concern is about the modeling of the observation noises using DPMs. Then, we focus on the suitability of this family of densities in the estimation problem handled by an improved particle filter called Rao-Blackwellized particle filter. The use of Rao-Blackwellization permits to

improve the efficiency of the filtering process since we do not need to compute everything with pure sampling which reduces significantly the computational cost. Another contribution is about the application context and the interesting validation schemes. The efficiency of the proposed approaches is demonstrated by applying a validation step involving simulated GNSS signals.

The remainder of this paper is organized as follows. Section II presents the pseudorange error modeling. Section III is dedicated to the Bayesian modeling of the problem. Section IV is about the nonparametric density estimation problem. We show how to tune the DPM hyperparameters in a flexible way for a better fitting of the data distribution shape. Section V discusses the use of an efficient Particle filter to perform optimal estimation. In section VI the efficiency of the proposed approaches is demonstrated by conducting validation experiments involving simulated GNSS signals. Section VII concludes this paper with a summary and discussion.

## II. PSEUDORANGE ERROR MODELING

In multi-sensors based systems, each sensor transmits a signal (an information) to the receiver (or antenna). In this paper, the GNSS (Global Navigation Satellite System) constellation [22] is considered as a sensor network. Consequently, each GNSS satellite is considered as a sensor [23]. These sensors work according to three different operation modes which are :

- Failure mode, when the sensor provide no information.
- Degraded mode, when the information delivered by the sensor is not accurate or is noisy.
- Normal mode, when the information delivered by the sensor can be considered as accurate.

The failure mode appears when the antenna cannot receive a signal from a satellite because of local masks. In the normal mode, the satellite signal reaches the antenna in LOS. Finally, the degraded mode occurs when the signal reaches the antenna after one or more reflections.

The navigation signal transmitted by each satellite includes a precise time at which the signal was transmitted. The distance or range from a receiver to each satellite may be determined using this time of transmission which is included in each navigation signal. By noting the time at which the signal was received at the receiver, a propagation time delay can be calculated. This time delay when multiplied by the speed of propagation of the signal will yield a *pseudorange* from the transmitting satellite to the receiver.

The range is called a pseudorange because the receiver clock may not be precisely synchronized to GPS time and because propagation through the atmosphere introduces delays into the navigation signal propagation times. These result, respectively, in a clock bias and an atmospheric bias. Clock biases may be as large as several milliseconds. Using information from at least four satellites, the position of a receiver with respect to the center of the Earth can be determined using triangulation techniques. The pseudorange measure deduced from the signal propagation time is expressed as follow :

$$\rho_t^s = d_t^s - c(\delta t_u + \delta t_s) + I_t^s + T_t^s + m_t^s + w_t \quad (1)$$

where

- $\rho_t^s$  is the pseudorange between the satellite  $s$  and the receiver at time  $t$ ;
- $d_t^s = \|p_t - p_t^s\|$  is the true satellite-receiver distance;  $p_t = [x_t, y_t, z_t]^T$  is the three position coordinates of

the vehicle in the reference system of coordinates East, North, Up (ENU)<sup>1</sup> and  $p_t^s = [x_t^s, y_t^s, z_t^s]^T$  is a vector containing the position coordinates of the satellite  $s$ ;

- $\delta t_u$  is the receiver clock offset with respect to the GPS reference time<sup>2</sup> and  $c = 3 \times 10^8$  m/s is the celerity (the speed of light);
- $\delta t_s$  is the satellite clock offset;
- $I_t^s$  and  $T_t^s$  are respectively the ionospheric and the tropospheric errors;
- $w_t$  is the receiver noise considered as a white Gaussian random variable with variance  $\sigma^2$  is general equal to unity;
- $m_t^s$  is the error due to the signal reflections in case of multipath propagations.

The atmospheric propagation errors and the satellites clock bias can be corrected by correction models [22]. Consequently, after corrections, (1) can be rewritten as (2):

$$\rho_t^s = d_t^s - c\delta t_u + m_t^s + w_t \quad (2)$$

It follows that the pseudorange error can be defined in function of two different error sources as :

$$\epsilon_t^s = m_t^s + w_t$$

where  $\epsilon_t^s$  is the sum of the error due to the signal reflections between the satellite  $s$  and the receiver and the receiver noise. According to the reception conditions,  $\epsilon_t^s$  can switch between different observation models as following :

$$\begin{cases} \text{LOS} & : m_t^s = 0, \epsilon_t^s \sim \mathcal{N}(0, \sigma^2) \\ \text{non-LOS} & : m_t^s \neq 0, \epsilon_t^s \sim \mathcal{N}(0, \sigma^2) \end{cases} \quad (3)$$

In the case when a signal is received in LOS, the pseudorange error distribution is considered white-Gaussian. In the other case, the pseudorange error distribution is unknown. In constrained environments, the density form can change abruptly or evolve slowly during a long period of time according to the obstacle nature. Moreover, several moving obstacles (vehicles, pedestrians, etc) can induce random errors. Therefore, to estimate accurately the pseudorange error density, a flexible model was proposed in this paper. It should be noted that our aim is to estimate at each time  $t$  the position  $p_t$  and the observation noise density  $m_t^s$  from the set of collected measurements  $\{\rho_t^{s_1}, \rho_t^{s_2}, \dots, \rho_t^{s_l}\}$  where  $l$  is the number of visible satellites at time  $t$ .

### III. BAYESIAN MODELING

In this section we are interested in designing a dynamic model for the parameters to be estimated. Our ultimate objective is to estimate accurately the state of a moving vehicle. We design by state the vector containing vehicle positions  $p_t = [x_t, y_t, z_t]^T$  and velocities  $\dot{p} = [\dot{x}_t, \dot{y}_t, \dot{z}_t]^T$ . Since we are interested in estimating jointly the hidden state  $\mathbf{x}_t = [p_t, \dot{p}_t]$  and the pseudorange noise density, we will consider an augmented hidden state vector containing both hidden state parameters and observation noise parameters. The observation error  $\epsilon_t$  is distributed according to

<sup>1</sup>This Cartesian coordinate system is far more intuitive and practical than ECEF or Geodetic coordinates [22]. The local ENU coordinates are formed from a plane tangent to the Earth's surface fixed to a specific location and hence it is sometimes known as a "Local Tangent" or "local geodetic" plane. By convention the east axis is labeled  $x$ , the north  $y$  and the up  $z$ .

<sup>2</sup>GPS has a reference time called GPS time maintained on the earth at the US naval observatory. The satellite clocks, though very accurate, are different pieces of equipment and so are not, in general, exactly synchronized with each other. Thus, they have different offsets with respect to GPS time.

an unknown probability density function  $F$  whose parameters denoted  $\theta_t$  is to be estimated. Let  $\mathcal{E}_t = [\mathbf{x}_t, \theta_t]$  stands for the vector containing all the parameters to be estimated.

In the following, a state-space approach providing a general framework for describing the dynamic state estimation problem at hand is provided. Here, we will use the state-space formulation to introduce the filtering algorithms. Statistical Bayesian filtering aims at computing the posterior probability density function of a state vector  $\mathbf{x}_t \in \mathcal{X}$  from sequentially obtained sensor measurements (pseudorange)  $\rho_t \in \mathcal{Z}$ . In our context, it relies on the following model :

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + v_t \quad (4)$$

$$\rho_t = h(\mathbf{x}_t, \epsilon_t) \quad (5)$$

$$\epsilon_t \sim F(\cdot) \quad (6)$$

where  $f$  is the state transition function,  $h$  is the observation function, they are assumed to be nonlinear, and  $v_t$  and  $\epsilon_t$  are respectively the state and the measurement noise vectors. Note that in this work we will be interested in estimating  $\epsilon_t$  and not  $v_t$ . We will assume that  $v_t$  is distributed according to a known density function with fixed parameters.

We are considering a real-world data analysis tasks which involves estimating unknown quantities from some given observations. In most of such applications, prior knowledge about the phenomenon being modeled is available. This knowledge allows us to formulate Bayesian models, that is prior distributions for the unknown quantities and likelihood functions relating these quantities to the observations. Within this setting, all inference on the unknown quantities is based on the posterior distribution obtained from the Bayes' theorem. Often, the observations arrive sequentially in time and one is interested in performing inference on-line. It is therefore necessary to update the posterior distribution as data become available. It is worth noting that computational simplicity in the form of not having to store all the data might be an additional motivating factor for sequential methods.

In this paper, we will focus on SMC integration methods [19] which are a set of simulation-based techniques that provide a convenient approach to computing the posterior distributions in our nonlinear and non-Gaussian case and lead most of the time to accurate filtering results. In the context of positioning with GPS, it is typically practical to process data on-line as it arrives from the visible satellites at each time  $t$ , both from the point of view of storage costs as well as for rapid adaptation to changing signal characteristics. In this paper the objective is to study a dynamic nonlinear model with additive observation errors. The inferential problem is to estimate sequentially the hidden states driving the observations along with the starting value. At the same time, our focus is to model the additive observation errors using a highly flexible family of density functions based on an infinite mixture model. In the following, in order to introduce the statistical Bayesian filtering, we first review some theoretical aspects about nonlinear dynamic state-space modeling. Then, we will provide some typical ways dealing with density estimation of errors via mixture based models.

#### A. Nonlinear dynamic state-space model and statistical Bayesian filtering

The unobserved signals  $\{\mathbf{x}_t; t \in \mathbb{N}\}$  are modeled as a *Markov Process* of initial distribution  $p(\mathbf{x}_0)$  and transition equation  $p(\mathbf{x}_t|\mathbf{x}_{t-1})$  provided by the motion model (4). The observations  $\{\rho_t; t \in \mathbb{N}^*\}$ , are assumed to be conditionally independent given the process  $\{\mathbf{x}_t; t \in \mathbb{N}\}$  and the marginal distribution  $p(\rho_t|\mathbf{x}_t)$  deduced from (5). To sum

up, the model is described by

$$\begin{cases} p(\mathbf{x}_0) \\ p(\mathbf{x}_t|\mathbf{x}_{t-1}), t \geq 1 \\ p(\rho_t|\mathbf{x}_t), t \geq 1 \end{cases}$$

We denote by  $\mathbf{x}_{0:t} \triangleq \{\mathbf{x}_0, \dots, \mathbf{x}_t\}$  and  $\rho_{1:t} \triangleq \{\rho_1, \dots, \rho_t\}$ , respectively, the states and the observations up to time  $t$ . Our aim is to estimate recursively in time the posterior distribution  $p(\mathbf{x}_{0:t}|\rho_{1:t})$  and its associated features (including the marginal distribution  $p(\mathbf{x}_t|\rho_{1:t})$ , known as the filtering distribution). The posterior distribution can be calculated in two steps theoretically: prediction and update. In the prediction step, we integrate the state distribution from the previous state using the system model. The update operation modifies the prediction distribution making use of the latest observation. Given the initial distribution  $p(\mathbf{x}_0)$ , transition distribution  $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ , and the likelihood distribution  $p(\rho_t|\mathbf{x}_t)$ , the objective of the filtering is to estimate the state at time  $t$  given the observations up to time  $t$ . From Bayesian perspective, when an observation  $\rho_t$  becomes available at time  $t$ , we can obtain this posterior distribution  $p(\mathbf{x}_{0:t}|\rho_{1:t})$  via the Bayes rule. The expression of this distribution requires the evaluation of complex high-dimensional integrals. Since we are in a nonlinear case, the multi-dimensional integrals are intractable and some approximate methods must be used. More details about the used filtering approach will be given in section V.

### B. Bayesian analysis of mixtures for density estimation

In this section, a Bayesian nonparametric approach based on a mixture model for the density estimation in a nonlinear dynamic model will be introduced. Later, we will study the inference procedure when the nonparametric component is applied to the additive observation errors  $\{\epsilon_t; t \geq 1\}$ . More precisely, we will show that by using a Dirichlet process prior [24], [13], [14] for the observation noise density  $F$ , we can derive a Bayesian nonparametric mixture model. Before we discuss the Dirichlet process prior and the nonparametric Bayes' analysis, let us first consider some typical ways to deal with a mixture model.

1) *General Framework*: Suppose we have observed data  $\rho_1, \dots, \rho_t$ , such that given  $(\mu_1, \Sigma_1), \dots, (\mu_t, \Sigma_t)$ , the  $\rho_i$ 's are independent, and the  $\rho_i$  has a normal distribution with mean  $\mu_i$  and variance  $\Sigma_i$ . Let  $\theta_i = (\mu_i, \Sigma_i)$ . Suppose a random variable  $\rho$  comes from any one from  $k$  components, each of them having a different probability density function. If  $\rho$  comes from  $i$ -th component with probability  $\pi_i$  and pdf  $f(\rho|\theta_i)$ , then the pdf of an observation  $\rho$  can be represented in the mixture form

$$f(\rho|\phi) = \sum_{i=1}^k \pi_i f(\rho|\theta_i), \quad (7)$$

where  $\phi = \{k, \pi_1, \dots, \pi_k, \theta_1, \dots, \theta_k\}$ ,  $k$  is the number of components, the  $\pi_i$ 's are the mixture weights which must be non-negative and sum to one and the  $\theta_i$ 's denote the vector of all unknown parameters associated with  $i$ th component. In some applications it is desired to categorize or classify each observation into one of the  $k$  populations. under a Bayesian approach this can be done by computing a classification probability, the probability that an observation comes a certain population. Alternatively, one may be interested in the elements of  $\phi$ , which include both the mixing weights  $\pi_1, \dots, \pi_k$ , and the density parameters  $\theta_1, \dots, \theta_k$ . The functional form of each distribution is usually assumed to be known. In some cases the number of components  $k$  may not be known, and is estimated from the data, as in cluster analysis [25]. Two kinds of Bayesian analysis of mixtures, parametric and nonparametric, have been developed. Parametric Bayesian analysis incorporates prior information about interesting



parameters in the inference. Sources of prior information include beliefs about parameters before any data are observed. On the other hand, Bayesian nonparametric mixture modeling has been developed recently. In this paper, our concern is about the nonparametric mixture models.

2) *Nonparametric Bayesian Analysis of Mixture Distributions:* Suppose data  $\rho_i (i = 1, \dots, t)$ , are distributed as  $f(\rho_i|\theta_i)$ , independently over  $i$ . The mixture structure simply imposes the constraint that, for some positive integer  $k$ , there exist  $k$  distinct numbers  $\theta^* = (\theta_1^*, \dots, \theta_k^*)$  such that, for each  $i = 1, \dots, t$ ,  $\theta_i = \theta_j^*$  for some  $j = 1, \dots, k$ . One way of generating such a mixture is to use a Dirichlet process for the prior distribution of  $\theta = (\theta_1, \dots, \theta_k)$ . An important result of the Dirichlet process is that the probability of  $\theta_i = \theta_j$  for  $i \neq j$  is positive. Such Bayesian analyses have been of limited use due to the computational difficulties. However, these difficulties can now be overcome using Gibbs sampling techniques [13]. This permits the extension of standard Bayesian parametric analyses to Bayesian nonparametric analyses, by linking the Gibbs sampling computations for Dirichlet process priors to those for the standard models.

#### IV. BAYESIAN NONPARAMETRIC DENSITY ESTIMATION

In [26] it was shown that an important characteristic of the observation error is the non-stationarity. In the context of dynamic models, the assumption of stationarity is wrong. In fact, when reflections occur; and by considering a long period of observation time, the pseudorange error distribution can be easily approximated by a finite gaussian mixture model (Fig. 1). However, by considering a short period of time, pseudorange errors can be modeled by only one gaussian distribution whose parameters (mean and variance) evolve over time. This drawback can be limited by the use of DPM. DPM models generate an infinite Gaussian mixture. In this mixture model the Dirichlet process prior is used to specify latent patterns of heterogeneity, particularly when the distribution of latent parameters is thought to be multimodal. It will shown through this paper that this method is flexible and well suited for on-line applications.

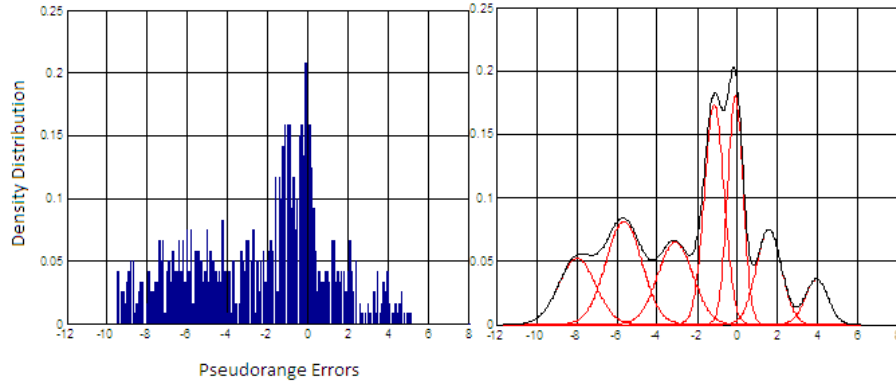


Fig. 1. The pseudorange errors can be approximated by a finite Gaussian mixture model where the number of components in the mixture should be specified in advance and the choice of the analysis window length should be done carefully by considering data evolution aspects.

In traditional application of mixture models, namely, the estimation of a (univariate) density function, the purpose is to determine the appropriate number of mixture components as well as to estimate their corresponding parameters. More generally, the problem of density estimation arises when data are considered to be sampled from a certain

distribution but this distribution is itself assumed unknown. Such problems (and their multiple variations) arise in nearly all branches of Statistics. Nonparametric Bayesian methods are specially well suited for density estimation. References can be found, among others in [27], [28], [29], [15]. The focus of our work will be on mixtures of normals. As it will be explained in this paper, the main motivation behind this choice is the application context.

#### A. Nonparametric Density estimation

In mixture models the Dirichlet Process is used to specify latent patterns of heterogeneity, particularly when the distribution of latent parameters is thought to be clustered. However, neither the number nor the location of clusters needs to be specified a priori. Therefore, DP mixture models are considered very attractive because DP mixtures allow for uncertainty in the choice of parametric forms and in the number of mixing components. The DP mixture model has found widespread application in recent statistical research [17], [30], [31], [32]. Analyses based on DP mixtures are sensitive to the choice of the Dirichlet process parameters; the need to treat these parameters carefully is discussed in [13].

Let consider a pdf  $F$  and  $\rho_1, \dots, \rho_t$  a set of vectors statistically distributed according to  $F$ ,  $\rho_k \sim F(\cdot)$  for  $1 \leq k \leq t$ . We will consider a nonparametric model allowing to estimate  $F$  as follows:

$$F(\rho) = \int_{\theta} f(\rho|\theta) d\mathbb{G}(\theta) \quad (8)$$

where  $\theta \in \Theta$  is called the latent variable (also called cluster),  $f(\cdot|\theta)$  is the mixed pdf and  $\mathbb{G}$  is the mixing distribution. Under a Bayesian framework, it is assumed that  $\mathbb{G}$  is a Random Probability Measure (RPM) distributed according to a prior distribution. In this paper, we have selected as a RPM the Dirichlet Process (DP) prior.

#### B. Dirichlet Processes

In [24] the Dirichlet Process (DP) has been introduced as a probability measure on probability measures. The DP is parameterized by a base distribution  $\mathbb{G}_0$  on a (measurable) space and a positive scaling parameter  $\alpha$ . Let's suppose now that we draw a random measure  $\mathbb{G}$  from a DP, and independently draw random variables  $\theta_k$  from  $\mathbb{G}$ :

$$\theta_k | \mathbb{G} \sim \mathbb{G} \quad (9)$$

where  $\mathbb{G} \sim DP(\mathbb{G}_0, \alpha)$ . Marginalizing out the random measure  $\mathbb{G}$ , the joint distribution of  $\{\theta_1, \dots, \theta_t\}$  follows a Polya urn scheme [33]. In another representation of the DP called the stick-breaking scheme,  $\mathbb{G}$  is introduced explicitly as an infinite sum of atomic measures [34]. The realizations of a DP are expressed as follow:

$$\mathbb{G} = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} \quad (10)$$

with  $\theta_k \sim \mathbb{G}_0$ ,  $\pi_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j)$  and  $\beta_k \sim \mathcal{B}(1, \alpha)$  where  $\mathcal{B}$  denotes the Beta distribution and  $\delta_{\theta_k}$  denotes the Dirac delta measure located in  $\theta_k$ . The underlying random measure  $\mathbb{G}$  is then discrete with probability one. Using (8), it comes that the following flexible prior model is adopted for the unknown distribution  $F$

$$F(y) = \sum_{k=1}^{\infty} \pi_k f(y|\theta_k) \quad (11)$$

More details about other important properties of Dirichlet processes can be found in [35], [33], [34], [14].

### C. Dirichlet Process Mixture (DPM)

DP mixture models form a very rich class of Bayesian nonparametric models. They emerge by employing a DP prior for the mixing distribution in a mixture of a parametric family of distributions. DP mixtures have dominated the Bayesian nonparametric literature after the machinery for their fitting, using Markov chain Monte Carlo (MCMC) methods, was developed following the work in [13].

DPM models [14] have become increasingly popular for modeling when conventional parametric models would impose unreasonably stiff constraints on the distributional assumptions. Recall that in the finite mixture model, each data point is drawn from one of fixed distributions. To allow the number of mixture components to grow with the data, we move to the general mixture model setting. This is best understood with the hierarchical graphical model in [13]. The core of the DPM model can basically be thought of as a simple Bayes model given by the likelihood  $y_k \sim f(y_k|\theta_k)$  and prior  $\theta_k \sim \mathbb{G}(\theta_k)$ , with added uncertainty about the prior distribution  $\mathbb{G}$  :

$$\begin{aligned} y_k|\theta_k &\sim f(y_k|\theta_k), \quad k = 1, \dots, t \\ \theta_k|\mathbb{G} &\sim \mathbb{G} \\ \mathbb{G} &\sim DP(\mathbb{G}_0, \alpha) \end{aligned}$$

Through this hierarchy, a marginalization step is often used in actual implementations. [33] proved that by integrating over  $\mathbb{G}$  the joint distribution of  $\theta = (\theta_1, \dots, \theta_t)$  may be factored into a product of successive conditional distributions of the following form :

$$\begin{aligned} \theta_1|\mathbb{G}_0 &\sim \mathbb{G}_0 \\ \theta_k|\theta_1, \dots, \theta_{k-1}, \alpha, \mathbb{G}_0 &\sim \frac{1}{k-1+\alpha} \sum_{j=1}^{k-1} \delta(\theta_j) + \frac{\alpha}{k-1+\alpha} \mathbb{G}_0, \quad \text{for } 2 \leq k \leq t \end{aligned} \tag{12}$$

This factorization implies that  $\theta$  is randomly partitioned into classes of distinct values such that the elements of  $\theta$  within a class share the same value. Therefore, a random draw from the Dirichlet process  $DP(\alpha, \mathbb{G}_0)$  may be computed as follows :

$$\theta_{t+1}|\theta_{1:t}, \mathbb{G} = \begin{cases} \theta_c^*, & \text{with probability } n_{t,c}/(\alpha + t) \\ \mathbb{G}_0, & \text{with probability } \alpha/(\alpha + t) \end{cases} \tag{13}$$

where  $\theta_c^* (c = 1, 2, \dots, k_c)$  denotes the  $c^{th}$  of  $k_c$  distinct values in  $\theta_{1:t} = (\theta_1, \dots, \theta_t)$  and  $n_{t,c}$  denotes the number of elements in  $\theta_{1:t}$  that equals  $\theta_c^*$ . In other words, a random draw from  $\mathbb{G}$  either equals one of the previous draws or is drawn independently from the base probability measure  $\mathbb{G}_0$ . This is due to the discreteness of the random measure  $\mathbb{G}$ .

### D. Estimation of DP Hyperparameters

In a hierarchical model, a hyperprior can be placed on the parameters of the DP by simply putting prior distributions on  $\alpha$  and the parameters of  $\mathbb{G}_0$ . The data is then used to calculate the posterior distribution. We can say that the DP adds another layer in the hierarchical model. The importance of treating the Dirichlet Process hyperparameters carefully is discussed in [13]. In applications, some authors have chosen a fixed  $\mathbb{G}_0$ , such as a normal distribution with large variance relative to the data [17] or a Beta( $a, b$ ), where  $a$  and  $b$  are stipulated without

further discussion [36]. Other authors have used a parametric family of priors for  $\mathbb{G}_0$ , such as uniform-inverse Wishart [37] or normal-inverse gamma [37], [38], with fixed hyperprior distributions on the parameters of the prior family. The hyperprior distributions are usually chosen by sensitivity analysis, or else they are chosen to be diffuse with respect to the observed data.

In this paper, Concerning the centering distribution  $\mathbb{G}_0$ , priors on  $\mathbb{G}_0$  hyperparameters will be specified using a data-adaptive way without applying a pure sampling approach. Most of the time, in real-world application contexts, the observed data can help in defining a prior for some of the hyperparameters in the considered model. Computing every unknown hyperparameter with pure sampling is not the most effective approach in all cases since this can be practically difficult and time consuming.

Since  $\alpha$  is an important parameter in this problem, one may want to put a prior on  $\alpha$  and then use the data to calculate a posterior distribution. In understanding what values of  $\alpha$  to use, the relationship between  $\alpha$  and the expected number of clusters in the data might be considered. In [13] was developed a clever Gibbs sampler wherein draws from the conditional posterior distribution of  $\alpha$  are computed by drawing successive samples from relatively familiar distributions (Beta and Gamma).

#### E. Prior and posterior distributions for $\alpha$

In [39], the author shows how, with respect to a flexible class of prior distributions for parameter  $\alpha$ , the posterior may be represented in a simple conditional form that is easily simulated. As a result, for a better fitting of DPM models, inference about this parameter may be developed with existing Gibbs sampling algorithms [13].

We have mentioned in this paper that a key feature of the Dirichlet is its discreteness, which in our context implies that the pairs  $(\mu_k, \Sigma_k)$ ,  $(k = 1, \dots, t)$ , concentrate on a set of some  $l \leq t$  distinct pairs. To sample the precision parameter  $\alpha$ , we first determine the prior distribution for  $l$ , the number of normal components in the mixture (clusters). At each stage of the simulation analysis, a specific value of  $l$  is simulated from the posterior for  $l$  (together with sampled values of the means and variances of the normal components) which also depends critically on this hyperparameter  $\alpha$ . In [13], [39] it was shown how, based on a specific but flexible family of prior distributions for  $\alpha$ , the parameter vector  $\theta$  may be augmented to allow for simulation of the full joint posterior now including  $\alpha$ .

As in [39], a Gamma prior for  $\alpha$  will be used. Suppose  $\alpha \sim G(a, b)$ , a gamma prior with shape  $a > 0$  and scale  $b > 0$ . In this case,  $p(l|\alpha, t)$  may be expressed as a mixture of two Gamma posteriors, and the conditional distribution of the mixing parameter given  $\alpha$  and  $l$  is a simple Beta.

$$(\alpha|x, l) \sim \pi_x G(a + l, b - \log(x)) + (1 - \pi_x) G(a + l - 1, b - \log(x)) \quad (14)$$

with weights  $\pi_x$  defined by

$$\frac{\pi_x}{(1 - \pi_x)} = \frac{(a + l - 1)}{t(b - \log(x))} \quad (15)$$

where  $0 < x < 1$ ,  $l > 1$  and  $(x|\alpha, l) \sim \beta(\alpha + 1, t)$ . More details about this sampling can be found in [39]. It is important to mention that  $\alpha$  should be sampled at each Gibbs iteration stage in the simulation process that allow to generate the  $\theta_k$  according to (18). The current sampled values of  $l$  and  $\alpha$  permit to compute a new value of  $\alpha$ .

In [35] the expected number of components sampled from a Dirichlet process is given by

$$\begin{aligned} E[l|\alpha, t] &= \sum_{l=1}^t l p(l|\alpha, t) \\ &= \alpha \sum_{l=1}^t \frac{1}{\alpha + l - 1} \\ &\approx \alpha \ln\left(\frac{t + \alpha}{\alpha}\right) \end{aligned} \quad (16)$$

Thus, although  $l \rightarrow \infty$  with probability 1 as  $t \rightarrow \infty$ , the number of components increases approximately logarithmically with the number of observations.

#### F. The $\mathbb{G}_0$ parameters

DP mixtures have received a considerable interest in the Bayesian nonparametric literature. Despite this activity, little has been written on choosing a centering distribution  $\mathbb{G}_0$  for these models. In this work, the latent sources of heterogeneity to be specified through the  $\mathbb{G}_0(\mu, \Sigma)$  parameters are considered using a data-adaptive way.  $\mathbb{G}_0$  is a normal distribution with unknown mean  $\mu$  and covariance  $\Sigma$ . The couple  $(\mu, \Sigma)$  can be chosen to be distributed according to a normal-scaled inverse Wishart distribution, denoted as :

$$(\mu, \Sigma) \sim \mathcal{NW}^{-1}(\lambda, \nu, \Sigma_0, dof) \quad (17)$$

such that :

$$\begin{aligned} \mu|\Sigma, \lambda, \nu &\sim \mathcal{N}(\lambda, \Sigma/\nu) \\ \Sigma|\Sigma_0, dof &\sim \mathcal{W}^{-1}(\Sigma_0, dof) \end{aligned}$$

where  $\lambda$  is the mean of the Normal distribution,  $\nu$  is a scale parameter,  $\Sigma_0$  and  $dof$  are respectively the inverse scale matrix and the degree of freedom for the inverse Wishart distribution  $\mathcal{W}^{-1}$  (see [14] for more details about these distributions).

The same base distribution  $\mathbb{G}_0$  should not be applied in all navigation situations (LOS, non-LOS and blocked). Here, we propose to adapt the priors on the  $\mathbb{G}_0$  parameters at each time  $t$ . More precisely, we propose to use an additional parameter  $r_t^s$  defined at each time  $t$  for each satellite  $s$ . This parameter adds one more step in the DPM hierarchical model as shown in Fig. 2.

The parameter  $r_t^s$  is an indicator of the propagation state for each satellite  $s$  at each time  $t$ . It will be used to update the parameters of  $\mathbb{G}_0$ . In practice, when a LOS reception state is detected the hyperparameters of the base distribution  $\mathbb{G}_0$  will be sampled from the normal-scaled inverse Wishart distribution whose parameters  $\lambda, \nu, \Sigma_0$  and  $dof$  should be chosen carefully. However, when a non-LOS reception state is detected, it will be interesting to adapt the distribution of the hyperparameters as following :

$$\begin{cases} \text{if } r_t^s = 0 \text{ (LOS)} & \text{then, } (\mu, \Sigma) \sim \mathcal{NW}^{-1}(\lambda_1, \nu_1, \Sigma_{01}, dof_1) \\ \text{if } r_t^s = 1 \text{ (non-LOS)} & \text{then, } (\mu, \Sigma) \sim \mathcal{NW}^{-1}(\lambda_2, \nu_2, \Sigma_{02}, dof_2), \end{cases} \quad (18)$$

with  $\Sigma_{02} > \Sigma_{01}$ ,  $dof_2 > dof_1$ ,  $k_2 < k_1$  and  $\lambda_2 = \lambda_1 + res_t^s$ . The additive term  $res_t^s$  is the estimation of the pseudorange error computed at each time  $t$  as following :

$$res_t^s = \rho_t^s - \|p_{t|t-1} - p_t^s\| - b_{t|t-1} \quad (19)$$

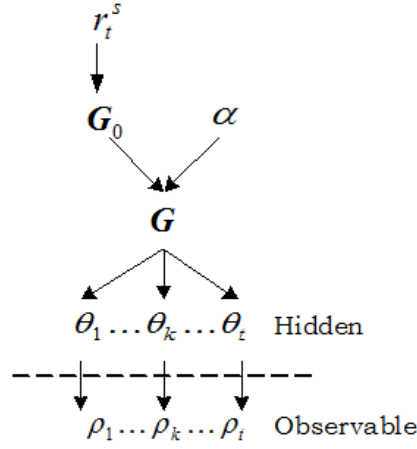


Fig. 2. DPM hierarchical model adaptation using the indicator  $r_t^s$  of the signal reception state.

where  $\rho_t^s$  is the corrected pseudorange measurement and  $p_{t|t-1} = [x_{t|t-1}, y_{t|t-1}, z_{t|t-1}]^T$  is the predicted position and  $b_{t|t-1} = c\delta t_u$  is the predicted receiver clock bias computed at time  $t$ .

## V. RAO-BLACKWELLIZED PARTICLE FILTERING FOR DYNAMIC BAYESIAN MODELS

### A. Sequential Monte Carlo Filtering

Unscented Kalman filters (UKF) and Extended Kalman filters (EKF) which are widely used as optimal solution in case of white Gaussian systems are not suitable for treating the models which exhibit non centered and non Gaussian distributions. Since, as we know, real data can be very complex. Except in a few simple cases, there is no closed-form solution to this problem. It is therefore necessary to adopt numerical techniques in order to compute reasonable approximations. Sequential Monte Carlo methods (particle filters) are powerful tools that allow us to accomplish this goal. This filtering method is able to cope with noises of any distribution and provide a convenient and attractive approach to computing the posterior distributions.

The particle filter approximates the posterior probability density function via the discrete weighted sum:

$$\hat{p}(x_t|\rho_{1:t}) \approx \sum_{i=1}^N \mathbf{w}_t^{(i)} \delta(x_t - x_t^{(i)}) \quad (20)$$

The individual weights  $\mathbf{w}_i$ , are computed by applying the principle of importance sampling. Since it is difficult or impossible to directly sample  $p(x_t|\rho_{1:t})$ , particles are simulated according to a proposal distribution :

$$x_t^{(i)} \sim q(x_t|x_{1:t-1}^{(i)}, \rho_{1:t}), \text{ for } i = 1, \dots, N$$

Then, to correct for the discrepancy between the proposal and the target distribution, importance weights will be assigned to the generated particles :

$$\mathbf{w}_t = \tilde{\mathbf{w}}_t^{(i)} / \sum_{k=1}^N \tilde{\mathbf{w}}_t^{(k)}$$

where

$$\tilde{\mathbf{w}}_t = \mathbf{w}_{t-1}^{(i)} \frac{p(\rho_t|x_{1:t-1}^{(i)}, \rho_{1:t-1})p(x_t^{(i)}|x_{1:t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, \rho_{1:t})} \quad (21)$$

According to (21), the set of particles is updated and reweighted using a recursive version of importance sampling. The most likely particles yield high importance weights. An additional resampling procedure is used for replacing particles with large weights and removing those with small weights. The variance introduced by the resampling procedure can be reduced by proper choice of the resampling method [19].

### B. Improving the Particle Filtering by Rao-Blackwellisation

A method to increase the efficiency of sampling techniques is to reduce the size of the state space by marginalizing out if possible some of the variables analytically; this is called Rao-Blackwellisation [21]. In the following, a Rao-Blackwellised particle filtering (RBPF) [40] will be introduced in order to improve filtering.

As showed previously, we can compute the posterior distribution on-line by applying Bayes' rule sequentially. However, in general, the required integrals cannot be computed in closed form. Particle filtering therefore approximates the posterior using sequential importance sampling. Now, if we partition the state-space into two sub-spaces, drawn by  $\theta_t$  and  $x_t$  (for all  $t \geq 1$ ). Then, by the chain rule of probability, we can write :

$$p(x_{0:t}, \theta_{1:t} | \rho_{1:t}) = p(x_{0:t} | \theta_{1:t}, \rho_{1:t}) p(\theta_{1:t} | \rho_{1:t}) \quad (22)$$

If we can update  $p(x_{0:t} | \theta_{1:t}, \rho_{1:t})$  analytically and efficiently, then we only need to sample  $p(\theta_{1:t} | \rho_{1:t})$  using the particle filter. Since we are now sampling in a smaller space, in general we will need far fewer particles to reach the same accuracy as standard PF. This is the key idea behind RBPF.

The following is a generalization of the Rao-Blackwellized particle filter to DPM as presented in [19] and applied in [41]. At time  $t$ , we use the following empirical distribution to approximate  $p(x_t | \rho_{1:t})$  through a set of  $N$  particles  $\theta_{1:t}^{(1)}, \dots, \theta_{1:t}^{(N)}$ ,

$$P_N(x_t | \rho_{1:t}) = \sum_{i=1}^N \tilde{w}_t^{(i)} p(x_t | \theta_{1:t}^{(i)}, \rho_{1:t}) \quad (23)$$

with

$$p(x_t | \theta_{1:t}^{(i)}, \rho_{1:t}) = \mathcal{N}(x_t; \hat{x}_{t|t}(\theta_{1:t}^{(i)}), \Sigma_{t|t}(\theta_{1:t}^{(i)})) \quad (24)$$

By using an Extended or an Unscented KF, for each particle  $i$  we can compute recursively the terms  $\hat{x}_{t|t}(\theta_{1:t}^{(i)})$  and  $\Sigma_{t|t}(\theta_{1:t}^{(i)})$ . For more details about how to build the algorithm refer to [41], [2]. In RBPF, each particle maintains not just a sample from  $p(\theta_{1:t} | \rho_{1:t})$  but also a parametric representation of the distribution  $p(x_t | \theta_{1:t}^{(i)}, \rho_{1:t})$  (the parametric representation in our case are a mean vector and a covariance matrix). For our application, the  $\theta_t$  samples are updated as in standard PF, and then the  $x_t$  distributions are updated using an EKF, conditional on  $\theta_t$ . The overall algorithm is shown in Algorithm 1.

Following what we presented above, the optimal importance distribution is  $q(\theta_t | \theta_{1:t-1}, \rho_{1:t}) = p(\theta_t | \theta_{1:t-1}, \rho_{1:t})$ . Without deriving approximations of this importance distribution, no efficient importance distributions can be used directly since the associated importance weights will be computationally intractable. In this paper, for the sake of simplicity, the importance distribution is chosen to be equal to the Polya urn distribution  $p(\theta_t | \theta_{1:t-1}^{(i)})$  as in (18). This approximation yields to the simplification of the weight expression (21).

Finally, the minimum mean squared error (MMSE) estimate and posterior covariance matrix of  $x_t$  are given by

$$\hat{x}_{t|t} = \sum_{i=1}^N \tilde{w}_t^{(i)} \hat{x}_{t|t}^{(i)} \quad (25)$$

$$\Sigma_{t|t} = \sum_{i=1}^N \tilde{w}_t^{(i)} [\Sigma_{t|t}^{(i)} + (\hat{x}_{t|t}^{(i)} - \hat{x}_{t|t})(\hat{x}_{t|t}^{(i)} - \hat{x}_{t|t})^T] \quad (26)$$

**Algorithm 1: Computing the posterior distribution  $p(x_{1:t}, \theta_{1:t} | \rho_{1:t})$  using a Rao-Blackwellized Particle Filter**

At time  $t = 0$ ,

**Step 0: initialization**

- For  $i = 1, \dots, N$ , sample  $(x_0^{(i)}, \Sigma_{0|0}^{(i)}) \sim p_0(x_0)$  and set  $t = 1$
- $\omega_0^{(i)} = \frac{1}{N}$

At time  $t \geq 1$ ,

**Step 1: Importance Sampling step**

- For  $i = 1, \dots, N$ 
  - Sample  $\tilde{\theta}_t^{(i)} \sim p(\theta_t | \theta_{1:t-1}^{(i)})$
  - Evaluate the importance weights  $\tilde{\omega}_t^{(i)} \propto \tilde{\omega}_{t-1}^{(i)} p(z_t | \theta_{1:t-1}^{(i)}, \tilde{\theta}_t^{(i)}, \rho_{1:t-1})$

**Step 2: Selection step**

- Multiply/Discard particles  $(\tilde{\theta}_t^{(i)}; i = 1, \dots, N)$  with respect to high/low normalized importance weights  $\tilde{\omega}_t^{(i)}$  to obtain  $N$  equally weighted particles.

**Step 3: Apply an Extended Kalman Filter to compute  $p(x_t | \tilde{\theta}_t^{(i)}, z_{1:t})$**

- Compute  $(x_{t|t-1}(\theta_{1:t}^{(i)}), \Sigma_{t|t-1}(\theta_{1:t}^{(i)}), x_{t|t}(\theta_{1:t}^{(i)}), \Sigma_{t|t}(\theta_{1:t}^{(i)}) = \text{EKF}(x_{t-1|t-1}(\theta_{1:t-1}^{(i)}), \Sigma_{t-1|t-1}(\theta_{1:t-1}^{(i)}), x_{t|t}(\theta_{1:t}^{(i)}), \Sigma_{t-1|t-1}(\theta_{1:t-1}^{(i)}, \tilde{\theta}_t^{(i)}, \rho_t))$

## VI. APPLICATIONS TO GNSS POSITIONING IN URBAN CANYON

### A. Error Density Modeling using DPM

For density estimation, the heterogeneity in the distribution of the latent parameters is a problematic issue in Bayesian nonparametric modeling. We know that in real applications, the sources of this heterogeneity cannot be completely known. We will prove through some experiments that the Dirichlet process prior may provide a satisfactory model in case where the sources of heterogeneity in the observation errors density are unobserved. The DP prior is robust to errors in model specification and allows heterogeneity in the patterns distributions to be specified in a data-adaptive way.

For example, the predictive distribution of densities based on our Dirichlet process prior can be multimodal (NLOS case) owing to its natural potential for identifying clusters. In contrast, standard parametric prior distributions of densities, such as the normal distribution, cannot capture multimodality because of their shape restrictions. But, if the actual distribution of density can be approximated by a unimodal distribution (LOS case), the Dirichlet process prior also adapts to this situation. In fact, the parametric model based on finite Gaussian mixtures is a limiting case of our modeling scheme when applying Dirichlet process prior with  $\alpha \rightarrow \infty$ . Therefore, we can say that the Dirichlet process prior is a natural extension of fully parametric models.

In our application of on-line pseudorange error density estimation, the latent variables  $\theta_i$  are the mean and the variance of each Gaussian distribution included in the infinite mixture, i.e.  $\theta_i = (\mu_i, \sigma_i)$ , for  $i \geq 1$ . To compute



a vehicle position at each step of the filtering process,  $N$  particles  $\theta_i^1, \dots, \theta_i^N$  are computed. The problem with GNSS applications is that signal propagation can be considered in different reception modes. In LOS reception, the pseudorange error noise follows a zero-mean Gaussian distribution with a standard deviation  $\sigma$  inferior or equal to 1 (actually it depends on the used receiver). In NLOS reception, the pseudorange error can be higher than 100 m. In the proposed DPM model, the  $\alpha$  will be estimated by sampling as detailed in section IV-D. The same base distribution  $\mathbb{G}_0$  will not be applied in all navigation situations. Here, we propose to adapt  $\alpha$  and  $\mathbb{G}_0$  each time using the proposed approaches in section IV-D.

---

**Algorithm 2: Tuning the hyperparameters of the DP**

For  $t = 1, \dots, n$

- For  $s = 1, \dots, N_s^t$  ( $N_s^t$ : Number of visible satellites at time  $t$ )
  - For each iteration  $i$  of the Gibbs sampler ( $i = 1, \dots, N$ ), do
    - \* Sample  $\alpha$  according to Algorithm 3
    - \* If LOS reception, set the following parameters of  $\mathbb{G}_0$ 

$$\mu_{t,s} \sim \mathcal{N}(\mu_0, \sigma_{t,s}/k_1)$$

$$\sigma_{t,s} \sim \mathcal{W}^{-1}(\Sigma_1, dof_1)$$
    - \* If NLOS reception, set the following parameters of  $\mathbb{G}_0$ 

$$\mu_{t,s} \sim \mathcal{N}(\mu_0 + res_t^s, \sigma_{t,s}/k_2)$$

$$\sigma_{t,s} \sim \mathcal{W}^{-1}(\Sigma_2, dof_2)$$

---



---

**Algorithm 3: Sampling the  $\alpha$**

For each iteration  $i$  of the Gibbs sampler ( $i = 1, \dots, N$ ), do

- Set the parameter  $\alpha$  to some predetermined value
- Consider  $n_t$  the number of measurements provided by the satellite  $s$  at time  $t$
- Compute the number of components in the mixture (clusters) at time  $t$  denoted  $k_t$ , ( $k_t \leq n_t$ )
- Set a prior for  $\alpha : \alpha \sim G(a, b)$  ( $a, b > 0$ )
- Compute a conditional distribution of the mixing parameter denoted  $\mathbf{x}$  using the Beta function as following
 
$$(\mathbf{x}|\alpha, k_t) \sim \beta(\alpha + 1, n_t)$$
- Use  $\mathbf{x}$  to compute weights  $\pi_{\mathbf{x}}$  as in (15)
- $p(\alpha|k_t, n_t)$  is a mixture of two gamma posteriors computed as in (14)

---

### B. Experimental set-up description

The Ergospace software simulates electromagnetic signal propagation in 3D realistic constrained environments. This software is developed to be a support for the development of GNSS applications. The deterministic method of Ray Tracing determines possible paths of received rays (example given in Fig. 3). Propagation models characterize interactions between the signal and the studied environment. Receiver behavior (mobile or other) and performance are also taken into account in calculations.

The software provides output data files in Matlab format that we process in order to compute the position. In this section, simulations have been performed with a 3D scene of the town of Rouen, in France. The simulated route

corresponds to an existing bus line. In these simulations, the mobile ran at a constant speed of 30 km/h, during 10 minutes and is covering a distance of 8236 meters. GNSS signals from both GPS and Galileo constellations can be received with a maximum of 3 reflections. In most urban transport applications, localization should be provided with a good accuracy. In our experiments, we have fixed the accuracy threshold to 3 meters [42].

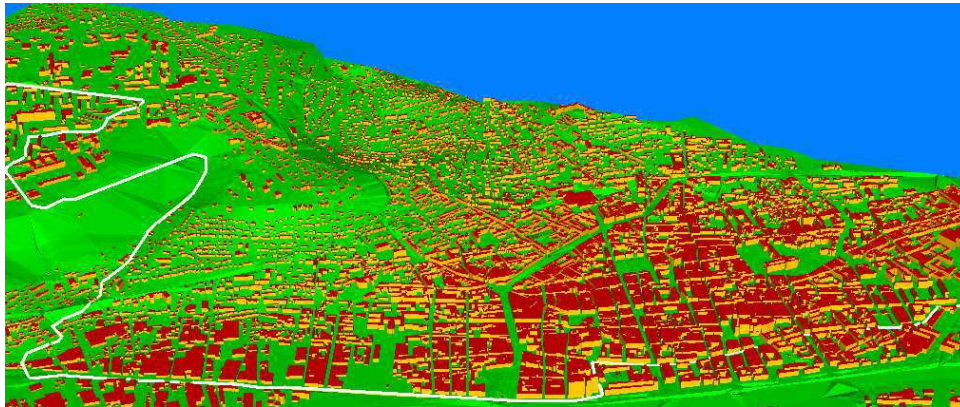


Fig. 3. An example illustrating the vehicle trajectory on a 3D realistic constrained environment (city of Rouen in France) using the Ergospace software simulator.

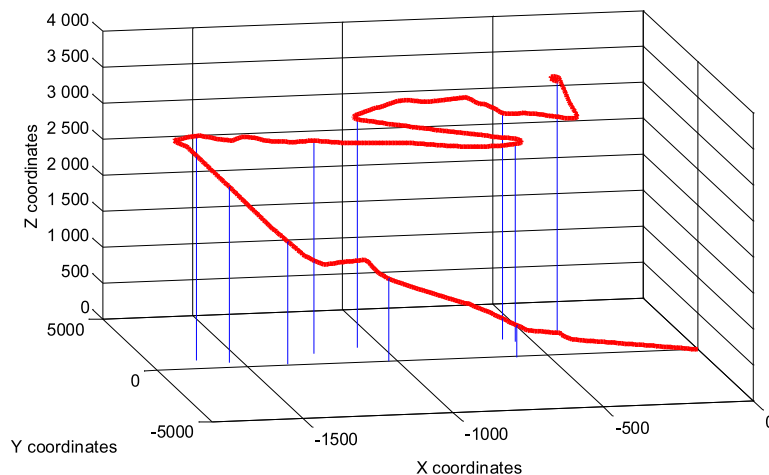


Fig. 4. The reference trajectory of a vehicle in an urban canyon: Positions are drawn in the local East North Up (ENU) Cartesian coordinate system. The unit of the vehicle coordinates is the meter (m).

### C. Results and discussion

In this part we will present some results that pertain to the choosing of the DPM model parameters. Moreover, we designed experiments involving some comparison with some commonly used estimation and filtering methods to show the robustness, flexibility and merits of our approach. Note that in the following, the DPM simulations are based on the implementation of the DP mixture model described in the previous subsection, applied to the case of conjugate models with normal structure. The precision parameter  $\alpha$  may be calibrated automatically following the

procedure proposed in Algorithm 3. The parameters of the centering distribution  $\mathbb{G}_0$  are simply adapted to the data characteristics as shown in Algorithm 2.

In order to highlight the effectiveness of the proposed algorithm based on RBPF, we have conducted preliminary experiments using EKF, PF and RBPF algorithms. Using these methods, localization performances were compared along the same trajectory during the same period of time. In the EKF algorithm, pseudorange error distributions are assumed to be centered Gaussian. In the PF algorithm, we considered a finite Gaussian mixture. About 500 particles are used in the above algorithms. Accuracy has been calculated with reference to an exact position given by the simulator (Fig. 4).

The use of RBPF algorithm is fully justified by the results presented below (Fig. 5 and Fig. 6), as it yields consistently lower error rates and a high localization accuracy. It appears that using DPM models can improve the accuracy and the availability of the localization even in very constrained environments.

Fig. 6 shows the evolution over time of the positioning error. The proposed algorithm has proven to be more accurate than the other algorithms. Besides, another key point permitting to conclude that those methods are more suitable than the others is the algorithm stability when abrupt propagation changes occur during mobile dynamic evolution (Fig. 5).

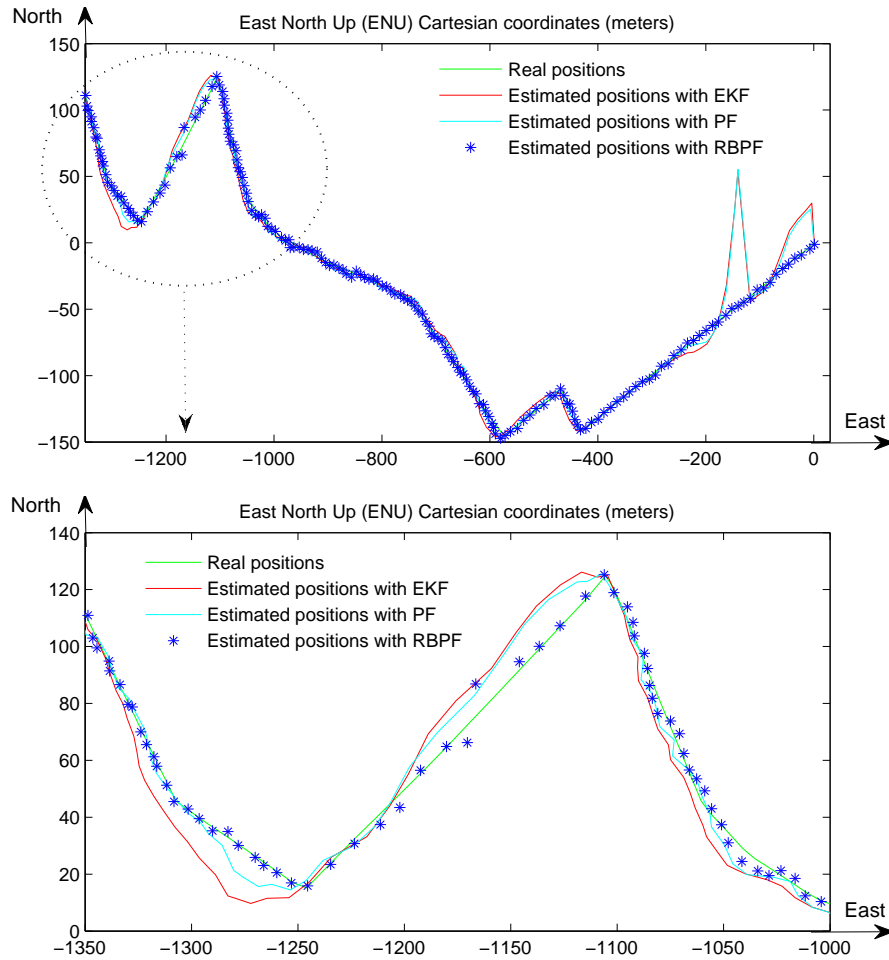


Fig. 5. Localization accuracy obtained by applying various filtering algorithms : EKF, PF (with GM for the density estimation) and the proposed approach based on RBPF (with DPM for the density estimation)

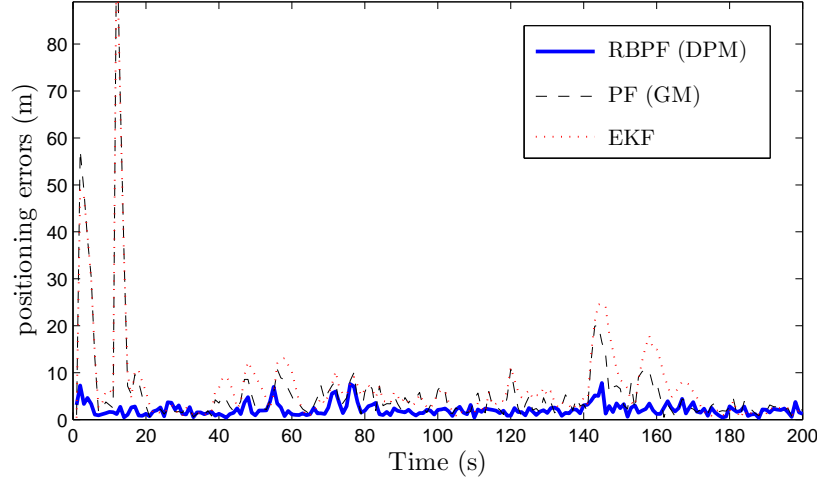


Fig. 6. Evolution of the positioning error for different tracking approaches : EKF, PF (GM is used for the density estimation of observation errors) and the proposed approach based on RBPF (DPM is used for density estimation of the errors)

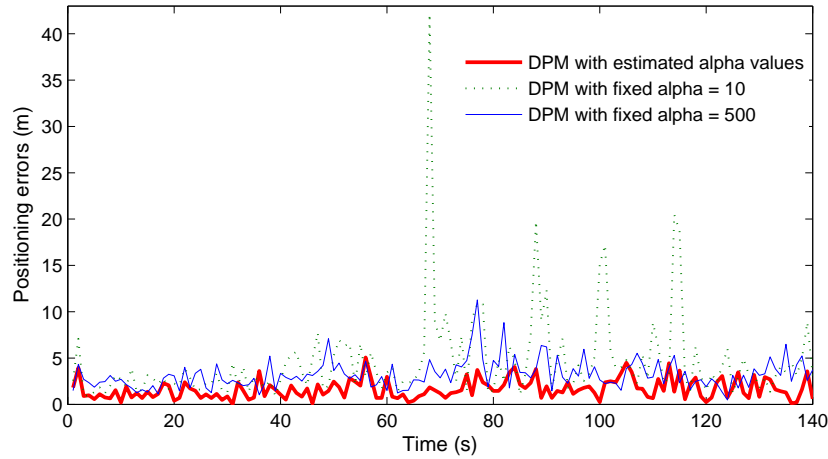


Fig. 7. Estimation of the pseudorange errors using DPM models with fixed and estimated values of  $\alpha$

The GM and the DPM are used in a sequential Monte Carlo filtering approach. Table I shows the localization performances for the two models in terms of mean error and availability of the navigation solution. An indicator of the availability of the navigation solution is a threshold over the error values which is set equal to 3 meters.

The parameters of a DP include the precision measure  $\alpha$  and the base probability measure  $\mathbb{G}_0$ . In problems where  $\alpha$  is unknown, estimation results can be extremely sensitive to the choice of prior assumed for  $\alpha$  [14]. In most applications of DPM models, the precision parameter  $\alpha$  is not known and must be well chosen. In this paper, some empirical tests were conducted to show the influence of  $\alpha$  values on the estimation performance.

In fact, a random draw from  $DP(\alpha, \mathbb{G}_0)$  either equals one of the previous draws or is drawn independently from the base probability measure  $\mathbb{G}_0$ . The parameter  $\alpha$  obviously plays an important role in the distribution of  $\theta$ . In equation (18), note that the probability that  $\theta_i$  differs from all previously drawn parameter values is proportional to  $\alpha$ ; therefore, higher values of  $\alpha$  lead to a higher probability of many unique values (i.e., a higher number of classes relative to the sample size  $n$ ). Conversely, lower values of  $\alpha$  lead to a greater chance of clustered distributions with

TABLE I  
COMPARISON OF POSITIONING ERRORS USING DPM AND GM

	Mean error	Min error	Max error	%error<3m
DPM	2.01	0.0014	8.284	89%
GM	6.14	0.0019	25.49	51%

fewer unique values in  $\theta$ .

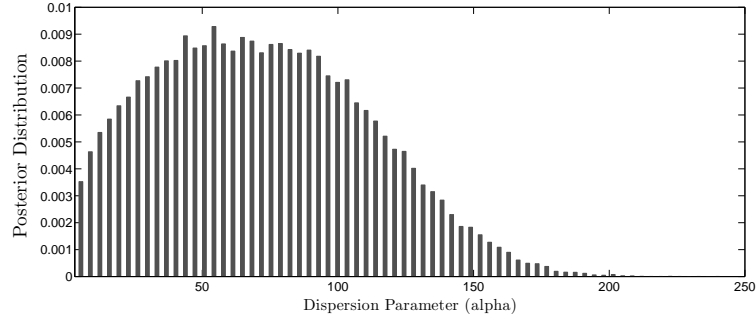


Fig. 8. Posterior distribution over the dispersion parameter  $\alpha$  when a Gamma prior is used.

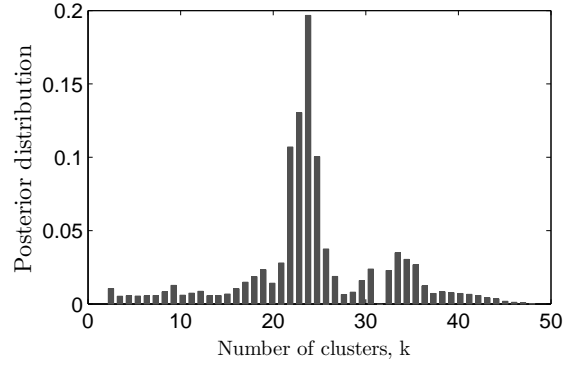


Fig. 9. Posterior distribution of the number of clusters  $k$

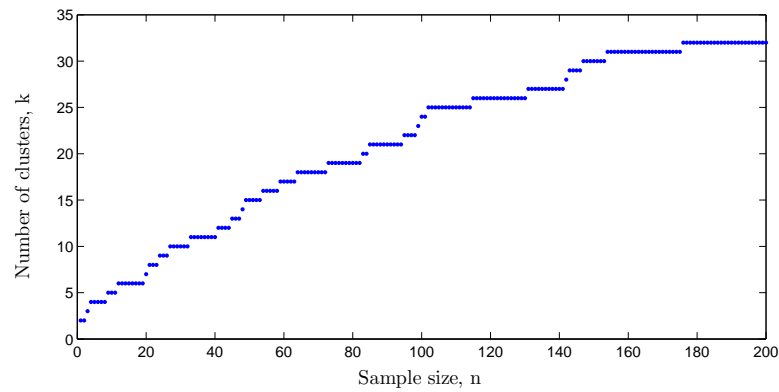


Fig. 10. The clusters number  $k$  evolves approximately logarithmically as new data come in.

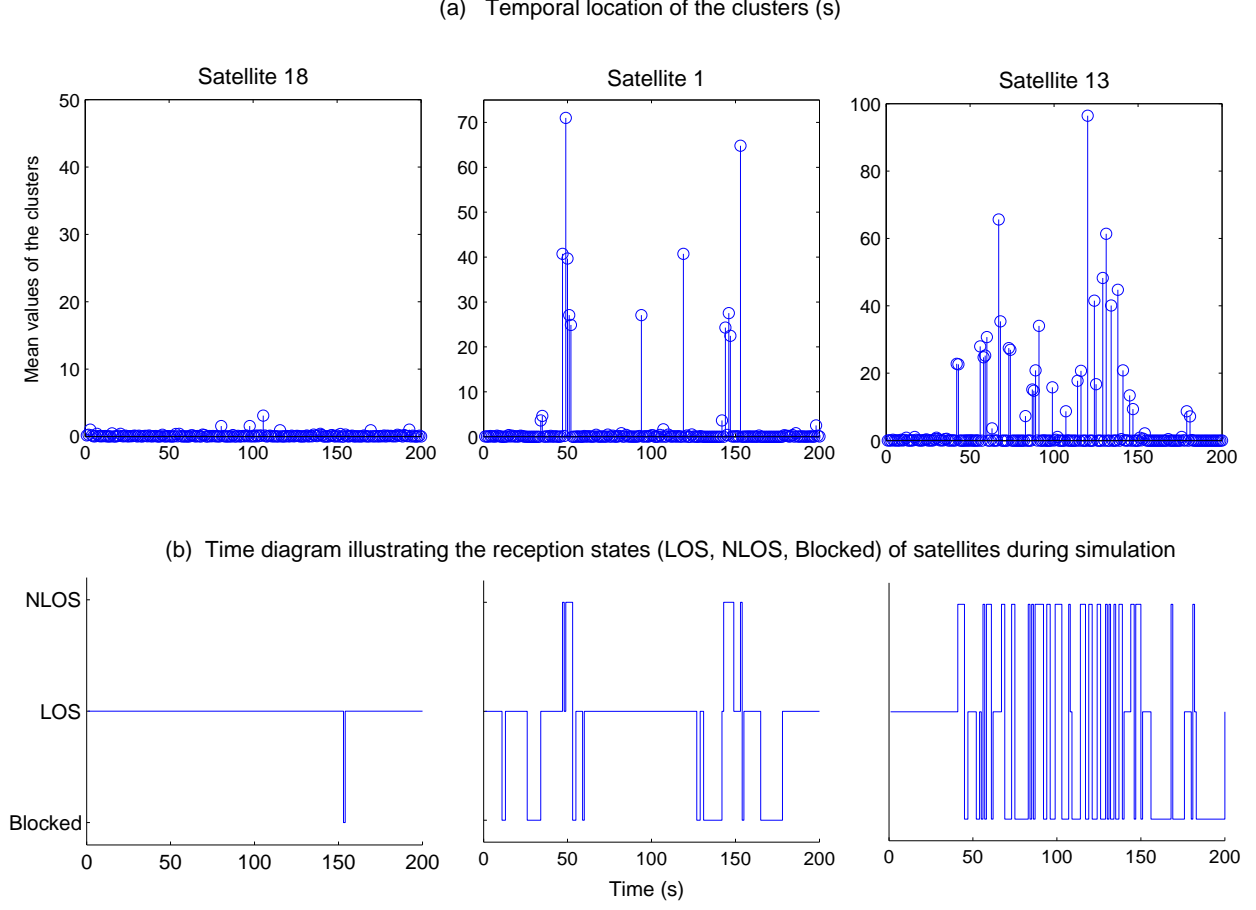


Fig. 11. Locations of the clusters in the DPM model for different satellites (satellites 18, 1 and 13) and their corresponding reception states.

In preliminary experiments, we considered a range of values in order to explore the extend to which our results are sensitive to different choices of  $\alpha$ . While the overall shape of the distribution remains the same across values of  $\alpha$ , higher values of  $\alpha$  appear to produce an increased number of localized features and even indicate additional modes for a subset of the observations. The estimated means and variances of the model parameters are, however, quantitatively very similar. Furthermore, we find little gain from using large values of  $\alpha$ . In our DPM model, we used a gamma distribution as a prior for the  $\alpha$ . The major advantage of this choice is that with this prior the conditional distributions are easy to sample by applying the approach of section IV-E. Fig. 7 shows that estimating the  $\alpha$  leads to a good positioning accuracy compared to the result obtained when we use a fixed  $\alpha$ .

The resulting posterior distribution over  $\alpha$  based on a gamma prior shown in Fig. 8 is considerably variable. Fig. 9 illustrates the posterior distribution of  $k$ . The posterior clearly indicates that  $k \ll n$ , which suggests that the distribution of the latent variables is clustered. Practically, in most situations a considerable uncertainty exists about the probable value of the number of model clusters  $k$ . In our context, nor the prior mean neither the prior variance of the  $k$  are known. As shown in (16), the number of clusters increases approximately logarithmically with the number of observations. This is illustrated in Fig. 10 which shows how the prior over the number of components changes as a function of  $n$ . Fig. 11 shows clusters and theirs locations for two satellites (signals from satellite 18

are mostly received in LOS and those of satellite 13 both in LOS and NLOS). It seems to be extremely hard to have such informations on data in a real-world application. Therefore, the parameters  $a$  et  $b$  of the gamma prior were adjusted carefully. For example, we noticed that small values of  $a$  and  $b$  lead to nearly similar values of the  $\alpha$  probability density, which means a lack of variability in the distribution of  $\theta$ .

## VII. CONCLUSION

Our work deals with the problem of improving localization accuracy in urban canyons when signals delivered by the GNSS system are received most of the time after many reflexions. The errors induced in the pseudorange measurements have an unknown density form. Thus, in this paper we have discussed a Bayesian nonparametric model based on Dirichlet process mixtures that permits robust adaptive inference to be carried out on data with some unknown specifications. The DP prior is extremely flexible, it allows to model a huge variety of distributional forms. In the considered application, its effectiveness was proved, since in our case little is known about priors to be considered for the errors distribution. We illustrated through experiments that a good choice of the DPM hyperparameters leads to a better density estimation. Since the proposed density estimation solution was strongly motivated by the complexity of the considered data, we proved that the proposed approaches outperform standard sequential methods, especially in a nonlinear dynamic context in which joint estimation of states and observation noise distributions is needed.

Many interesting questions remains to be addressed. One of these is whether non-stationarity can be involved in the nonparametric model by considering time-varying Dirichlet process mixtures. At each time step it can be interesting to consider that the unknown distribution follows a DPM model. Some interesting and simple approaches, such as utilized in [43], need to be investigated. A second is to develop more accurate and efficient approximations of the importance distribution  $q(\theta_t | \theta_{1:t-1}^{(i)}, z_{1:t})$  in the proposed RBPF.

## REFERENCES

- [1] A. Giremus, J.-Y. Tournet, and V. Calmettes, "A particle filter approach for joint detection/estimation of multipath effects on gps measurements," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1275–1285, 2007.
- [2] F. Caron, M. Davy, and E. Duflos, "Particle filtering for multisensor data fusion with switching observation models. Application to land vehicle positioning," in *IEEE transactions on Signal Processing*, 2006.
- [3] F. D. Nahimana, E. Duflos, and J. Marais, "Reception state estimation of gnss satellites in urban environment using particle filtering," in *FUSION*, Cologne, Germany.
- [4] N. Viandier, D. Nahimana, J. Marais, and E. Duflos, "GNSS performance enhancement in urban environment based on pseudo-range error model," in *IEEE/ION Position, Location and Navigation Symposium*, 2008, pp. 377–382.
- [5] Y. Bar-Shalom and X. R. Li, *Multitarget-multisensor tracking: principles and techniques*. YBS publishing, 1995.
- [6] J. Tugnait, "Adaptive estimation and identification for discrete systems with markov jump parameters," *IEEE Transactions on Automatic Control*, vol. 27, no. 5, pp. 1054–1065, 1982.
- [7] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [8] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: a survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103–123, 1998.
- [9] C. Andrieu, M. Davy, and A. Doucet, "Efficient particle filtering for jump markov systems. application to time-varying autoregressions," *IEEE Transactions on signal processing*, vol. 51, no. 7, 2003.
- [10] A. Doucet and B. Ristic, "Recursive state estimation for multiple switching models with unknown transition probabilities," *IEEE Transactions on Automatic Control*, vol. 38, no. 3, pp. 1098–1104, 2002.

- [11] V. Jilkov and X. R. Li, "Online bayesian estimation of transition probabilities for markovian jump systems," *IEEE Transactions on signal processing*, vol. 52, no. 6, pp. 1620–1630, 2004.
- [12] A. Doucet, N. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump markov linear systems," *IEEE Transactions on signal processing*, vol. 49, no. 3, pp. 613–624, 2001.
- [13] M. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, pp. 577–588, 1994.
- [14] R. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 9, pp. 249–265, 2000.
- [15] P. Muller and F. Quintana, *Nonparametric Bayesian data analysis*. statistical science 19, 2004.
- [16] A. E. Gelfand, A. Kottas, and S. MacEachern, "Bayesian nonparametric Spatial Modeling with Dirichlet process mixture mixing," *Journal of the American Statistical Association*, vol. 100, pp. 1021–1035, 2005.
- [17] H. Ishwaran and L. F. James, "Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information," *Journal of Computational and Graphical Statistics*, vol. 11, pp. 508–532, 2002.
- [18] R. M. Dorazio, "On selecting a prior for the precision parameter of Dirichlet process mixture models," *Journal of Statistical Planning and Inference*, pp. 3384–3390, 2009.
- [19] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer, 2001.
- [20] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter*. Artech House, 2004.
- [21] G. Casella and C. Robert, "Rao-Blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.
- [22] E. Kaplan and C. Hegarty, *understanding GPS: Principles and Applications*. Artech House, 2005.
- [23] A. Rabaoui, N. Viandier, J. Marais, and E. Duflos, "Studies on dpm for the density estimation of pseudorange noises and evaluations on real data," in *IEEE/ION Position Location and Navigation Symposium*, 2010, pp. 377–382.
- [24] T. Ferguson, "A Bayesian analysis of some nonparametric problems," *The annals of statistics*, vol. 1, pp. 209–230, 1973.
- [25] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings," *Journal of the American Statistical Association*, vol. 78, pp. 553–584, 1983.
- [26] N. Viandier, A. Rabaoui, J. Marais, and E. Duflos, "Gnss pseudorange error density tracking using dirichlet process mixture," in *13th International Conference on Information Fusion (FUSION)*, 2010.
- [27] E. Fox, E. Sudderth, and A. Willsky, "Hierarchical dirichlet processes for tracking manoeuvring targets," in *In Proceedings of the International Conference on Information Fusion*, 2007.
- [28] Y. W. Teh and M. I. Jordan, "Hierarchical Bayesian nonparametric models with applications," in *Bayesian Nonparametrics: Principles and Practice*. Cambridge University Press, 2010.
- [29] T. L. Griffiths and Z. Ghahramani, "Infinite latent feature models and the indian buffet process," in *Advances in Neural Information Processing Systems (NIPS)*, 2006.
- [30] E. R. Brown and J. G. Ibrahim, "A Bayesian semiparametric joint hierarchical model for longitudinal and survival data," *Biometrika*, vol. 59, pp. 221–228, 2003.
- [31] M. D. Iorio, P. Muller, G. L. Rosner, and S. N. MacEachern, "An ANOVA model for dependent random measures. Journal of the American Statistical Association," *Biometrika*, vol. 99, pp. 205–215, 2004.
- [32] P. Muller, F. Quintana, and G. L. Rosner, "A method for combining inference across related nonparametric Bayesian models," *Journal of the Royal Statistical Society Series B*, vol. 66, pp. 735–749, 2004.
- [33] D. Blackwell and J. MacQueen, "Ferguson distributions via Polya urn schemes," *The Annals of Statistics*, vol. 1, no. 2, pp. 353–355, 1973.
- [34] J. Stthuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.
- [35] C. E. Antoniak, "Mixtures of dirichet processes with applications to bayesian nonparametric problems," *Annals of Statistics*, vol. 2, pp. 1152–1174, 1974.
- [36] J. S. Liu, "Nonparametric hierarchical Bayes via sequential imputations," *Annals of Statistics*, vol. 24, pp. 911–930, 1996.
- [37] S. N. MacEachern and P. Muller, "Estimating mixture of Dirichlet process models," *Journal of Computational and Graphical Statistics*, vol. 7, pp. 223–238, 1998.
- [38] C. E. R. an Z. Ghahramani, "The infinite gaussian mixture model," in *NIPS*, 2000, pp. 294–300.
- [39] M. West, "Hyperparameter estimation in dirichlet process mixture models," Duke University, Tech. Rep., 1992.
- [40] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics ans Computing*, vol. 10, no. 3, pp. 197–208, 2000.



- [41] F. Caron, M. Davy, A. Doucet, and E. Duflos, “Bayesian inference for dynamic models with dirichlet process mixtures,” in *International Conference on Information Fusion*, 2006, pp. 10–13.
- [42] N. Viandier, J. Marais, A. Prestail, and E. D. Verdalle, “Positioning urban buses: GNSS performances,” in *ITST*, 2008, pp. 51–55.
- [43] F. Caron, M. Davy, and A. Doucet, “Generalized polya urn for time-varying dirichlet process mixtures,” in *23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2007.